

Exercise 2

(1) und (2) contain the English and German texts we are working with in the exercises for the course.

- (1) At the age of one, Harry had somehow survived a curse from the greatest dark sorcerer of all time, Lord Voldemort, whose name most witches and wizards still feared to speak. Harry's parents had died in Voldemort's attack, but Harry had escaped with his lightning scar, and somehow — nobody understood why — Voldemort's powers had been destroyed the instant he had failed to kill Harry.

[J.K. Rowling, *Harry Potter and the Chamber of Secrets*]

- (2) Damals im Alter von einem Jahr, überlebte Harry auf merkwürdige Weise den Todesfluch des größten schwarzen Magiers aller Zeiten. Die meisten Hexen und Zauberer hatten immer noch Angst, dessen Namen auszusprechen: Lord Voldemort. Harrys Eltern starben bei Voldemorts Überfall, doch Harry kam mit der blitzförmigen Narbe davon. Voldemorts Macht jedoch fiel in eben jenem Augenblick in sich zusammen, als es ihm mißlungen war, Harry zu töten. Und keiner konnte das begreifen.

[J.K. Rowling, *Harry Potter und die Kammer des Schreckens*]

1 Chunk Parsing

Instead of doing deep parsing of sentences, many applications just need to have rough information about sentence *chunks*. More often, just having information about NP-chunks is sufficient.

1. Identify NP-chunks in whichever of the Harry Potter excerpt you have chosen to work on.
2. How does your result compare with the output of a real system? Use one of the on-line chunking systems that are available.

2 Deep Parsing with large LFG Grammars

1. Now try parsing whichever text you have chosen to work with via one of the large-scale LFG grammars via the XLE Web Interface (<http://iness.uib.no/xle-web/xle-web>).

Hints:

- If a sentence does not work, try cutting it up into smaller parts and parsing these parts first. Then successively add more material until the sentence stops working — this should help you identify what works and what doesn't.
 - In the column labeled “Discriminants” you can disambiguate between various options. For example, if you try parsing *The monkey saw the bird in the tree.*, you can choose between the reading where *in the tree* modifies *bird* vs. the verb.
 - Note that some of the longer sentences might take quite a while to be parsed.
 - Also note that the grammar have been written and tested mainly on newspaper texts rather than literature. This will probably become obvious quite quickly.
2. Can all the sentences be parsed? If not, can you figure out why not? (sometimes the machine may also just “time out” when the parsing takes too long).
 3. How many solutions do the sentences (or the partial sentences) receive?
 4. What do you judge the quality of the resulting parses to be? Provide some discussion of the overall results.

3 Language Complexity

1. Which of these formal expressions and language examples illustrate which parts of the Chomsky Hierarchy? Note that for the purpose of this exercise you do not need to distinguish between *mildly context-sensitive* and *context-sensitive*. The following parts are relevant for this exercise: *regular language*, *context-free* and *context-sensitive*.

- Phrase Structure Rule: $S \rightarrow NP VP$
 - Morphological Analysis: $slept \leftrightarrow sleep + Verb + Past$
 - Annotated Phrase Structure Rule: $VP \rightarrow V NP: (\uparrow OBJ) = \downarrow$
2. Why can $a^n b^n$ not be a regular language?
 3. Which examples from which language show that natural language is not a regular language and why?

4 Lexical Semantics

Now let's look at the lexical semantics of words in our texts. Use the on-line versions of WordNet and the Unified Verb Lexicon (VerbNet, PropBank, FrameNet) for this exercise.

1. For some purposes it might be interesting to know that nouns like *instant* are a temporal noun. From which resource could this information be extracted and in what form?
2. The particular type of the PP is also often of interest. Could you figure out via WordNet's Hypernym relation that *attack* in *in Voldemort's attack* is an event/activity, but that *castle* (*in Voldemort's castle*) is a concrete entity? If so, how is the information encoded?
3. It would also be good to be able to figure out automatically that *kill*, *die* and *destroy* are related in terms of their lexical semantics. Explore WordNet and check whether/how one can establish a relationship between these two verbs.
4. Now let us look at the verb *fail*. This verb expresses that something did not succeed. Can you extract this information from the Unified Verb Index *ziehen*? (Hint: find *fail* and then look up *succeed*).
5. Now look at the verbs *survive* and *die*. Do the various lexical resources of VerbNet, PropBank or FrameNet postulate a direct connection between these verbs? If not, should they? Can I rely on the lexical resources if I am trying to find out how many arguments these two verbs each have? Why or why not (discuss each of the resources individually).

5 Machine Translation

Translate the file Exercise Uni KN.docx from English into German (you can also select another target language. In that case you won't be able to use the Translation Memory that we used in class) using the web-based CAT-Tool MemSource (as shown in class).

Use the Translation Memory and the terminology database that you used in class. In the process of the translation you should explore some additional functionalities that were not shown in class (and/or consult the online help), so that you can answer the following questions:

1. How can you perform a quantitative analysis of the text, indicating how many matches and repetitions you can expect during the translation?
2. How can a term be added to the termbase during translation?
3. How can you split or join segments that are too long or too short, respectively?

Please hand in the following files:

- Your translation — the translation will only be checked for completeness and consistency not for linguistic quality.
- The answers to the questions above.
- The username and the name of the Translation memory that you worked with.

In case you get stuck completely, please contact Martin Kappus at kapm@zhaw.ch.